

Emoción, Percepción, Producción: Un Estudio Psicolingüístico para Detectar Emociones en el Habla

Trabajo de Fin de Grado



Facultad de Filosofía y Letras
Grado de Filología Hispánica
Curso 2019-2020

Autor: Mónica González Machorro
Tutor: Mark Gibson

ÍNDICE

1. Resumen y palabras clave.....	5
1.1 Abstract and key words.....	5
2. Introducción.....	6
2.1 Aspectos psicológicos y fisiológicos de las emociones: feliz y triste..	6
2.2 Producción del habla.....	7
2.3 Las emociones en el habla	8
2.4 Percepción de las emociones en el habla	11
3. Proyecto 1	12
3.1 Experimentos: análisis acústico y percepción acústica.....	12
3.2 Resultados: análisis acústico	13
3.2 Resultados: percepción acústica.....	14
3.3 Conclusión y discusión	15
4. Proyecto II.....	16
4.1 <i>Masked priming</i> en el lenguaje	16
4.2 Experimento	17
4.3 Metodología	17
4.4 Análisis estadístico.....	18
4.5 Resultados	18
4.5.1 Resultados hispanohablantes.....	19
4.5.2 Resultados anglohablantes	20
4.6 Conclusión	21
5. Relevancia y futuros estudios.....	22
6. Bibliografía	23

1. RESUMEN Y PALABRAS CLAVE.

Este trabajo se enfoca en la detección de las emociones en el habla mediante un análisis acústico y una técnica psicolingüística conocida como *priming*. Las emociones estudiadas son: felicidad, tristeza y el estado neutro. Este estudio está dividido en dos proyectos. El proyecto 1 consiste en documentar los parámetros acústicos de las tres emociones e investigar, a través de un experimento de percepción, si los oyentes utilizan estos parámetros para detectar las emociones. Se examinaron las respuestas de hispanohablantes y anglohablantes. El proyecto 2 tiene como objetivo entender la interacción entre los parámetros acústicos detectados en el primer proyecto y el procesamiento semántico. Este estudio pretende contribuir en la investigación en el campo del reconocimiento de las emociones en la voz y su aplicación en el desarrollo de sistemas de inteligencia artificial.

Palabras clave: *emociones, habla afectiva, fonética, psicolingüística, procesamiento el habla, percepción, reconocimiento de emociones*

1.1 Abstract and key words

This study centers on the detection of emotions in speech using an acoustic analysis and a technique known as masked priming. Three emotions were studied: happiness, sadness and a neutral state. The study is divided into two main projects. Project 1 examines the different acoustic parameters found in the auditory stimuli and whether the subjects are able to identify these emotions in isolated, non-emotionally charged conditions. The second project aims to understand the relationship between the acoustic parameters and the semantic processing by using a masked priming technique in incongruent and congruent contexts. Results show that non-native speakers are more susceptible to priming effects due to differences in the interactions of visual/auditory modalities, and subsequent semantic processing. We situate the relevance of the study within the broader context of the study of speech emotion recognition and its application in machine learning algorithms and other models using deep neural networks

Key words: *emotions, emotional speech, phonetics, psycholinguistics, speech processing, perception, emotion recognition.*

2. INTRODUCCIÓN

Este trabajo se centra en la percepción y producción de emociones en el habla. Las emociones producen cambios fisiológicos que repercuten en los distintos parámetros acústicos, los cuales pueden ser percibidos por oyentes. Esta investigación se compone de dos proyectos. El primero consiste en un análisis acústico de diferentes estímulos acústicos y un experimento de percepción. El segundo proyecto busca entender cómo se integra la modalidad acústica y semántica en el habla afectiva. Al final propondremos diferentes vías en las que se puede continuar el estudio de esta área.

Entender qué parámetros son importantes para percibir el habla de las personas que nos rodean es un estudio complejo y variado, debido a que existen muchos factores involucrados en la comunicación verbal y no verbal, por ejemplo, los movimientos corporales y faciales (kinésica), el tono de voz (fonética), el contenido lingüístico, el sentido del mensaje (pragmática).

Charles Darwin, uno de los primeros científicos en estudiar las emociones tanto en seres humanos como animales, apuntó que, a pesar de que el ser humano cuenta con un conocimiento intuitivo sobre las experiencias humanas, este estudio es difícil “debido a que los movimientos son a menudo sutiles. Puede percibirse con nitidez un matiz distinto y, no obstante, resultar imposible precisar en qué consiste esa diferencia.” (Darwin 44). A partir de Darwin, la investigación en la expresión de las emociones incrementó exponencialmente, especialmente en el ámbito de las expresiones faciales.

Fue hasta el auge en el estudio del procesamiento de las lenguas aplicadas al aprendizaje automatizado —*machine learning*— cuando el estudio de la expresión de las emociones en la voz comenzó a alcanzar un desarrollo acelerado. Numerosos estudios (Scherer; Johnstone; Bachorowski/Owren; Bachorowski; Menezes/Erickson/Han; Laukka) han concluido que es posible detectar las emociones a través de la voz tanto en contextos lingüísticamente similares como distintos. Es decir, la expresión de las emociones en la voz no depende de mecanismos propios de un idioma, y, por lo tanto, su detección tiene la capacidad de universalizarse. Es importante señalar que estos resultados no han sido lo suficientemente conclusos como los obtenidos en estudios sobre las expresiones faciales, debido a la falta de unanimidad en el corpus utilizado (Johnstone).

La producción y percepción del habla afectiva es un campo plenamente interdisciplinario, debido a que requiere conocimientos de programación, por ejemplo, para programar experimentos de percepción; conocimientos matemáticos para realizar un análisis estadístico; conocimientos psicológicos para comprender el procesamiento cognitivo del habla; y lingüísticos con el fin de entender cómo se produce el código. Además, este estudio se complica, ya que los efectos de las emociones en la voz aparecen sobrepuestos al código lingüístico, y es el papel del investigador de poder distinguir ambos sin perjudicar uno de ellos (Johnstone).

2.1 Aspectos psicológicos y fisiológicos de las emociones: feliz y triste

Las emociones son acciones motivadas fisiológicamente y neuroquímicamente que representan una reacción a eventos de gran significado para un individuo (Scherer). Es una “experiencia multidimensional con al menos tres sistemas de respuesta: cognitivo,

expresivo y fisiológico” (Chóliz 3). Una emoción activa el sistema nervioso autonómico, el cual cambia el flujo de las secreciones glandulares y neurohormonales (Izard).

El sistema límbico en el cerebro emite hormonas como serotonina y dopamina a través de neurotransmisores (Simonyan/Horwitz/Jarvis). Estos procesos fisiológicos y neuroquímicos también afectan al habla. Por ejemplo, con respecto a la dopamina — neurotransmisor responsable de la felicidad— se ha encontrado que en exceso produce espasmos en la laringe (Simonyan/Horwitz/Jarvis). En un paciente con Parkinson, la pérdida de neuronas dopaminérgicas lleva a la rigidez muscular y temblores, esto se manifiesta en un habla con una mayor aperiodicidad. En un estudio (Hornykiewicz), 90% de los pacientes con Parkinson también presentaron cambios significativos en la calidad de la voz, caracterizados por un habla monótona y un bajo tono de voz.

Las emociones básicas, señaladas por el Dr. Ekman, son: alegría, tristeza, asco, miedo, ira y sorpresa (Chóliz 7). Se puede hablar de emociones básicas ya que “existe una serie de universales en la expresión emocional demostrados transculturalmente, así como por un patrón fisiológico que caracteriza a cada una de ellas.” (Chóliz 8). En este trabajo nos vamos a enfocar en dos emociones: alegría y tristeza. Además, tendremos en cuenta un estado neutro, debido a que todas las “emociones básicas tienen un sustrato neutral distintivo”. (Chóliz 7).

La alegría se define, por lo tanto, como un proceso que “favorece la recepción e interpretación positiva de los diversos estímulos ambientales” (Chóliz 10). Esta emoción se caracteriza por un “aumento en la actividad del hipotálamo, septum y núcleo amigdalino; aumento en la frecuencia cardíaca e incremento en la frecuencia respiratoria” (Chóliz 10). La tristeza se relaciona con una “actividad neurológica elevada y un ligero aumento en frecuencia cardíaca y presión sanguínea” (Chóliz 16). Estas características neurológicas se manifiestan en el habla de diversas formas, parámetros que serán explicados a continuación.

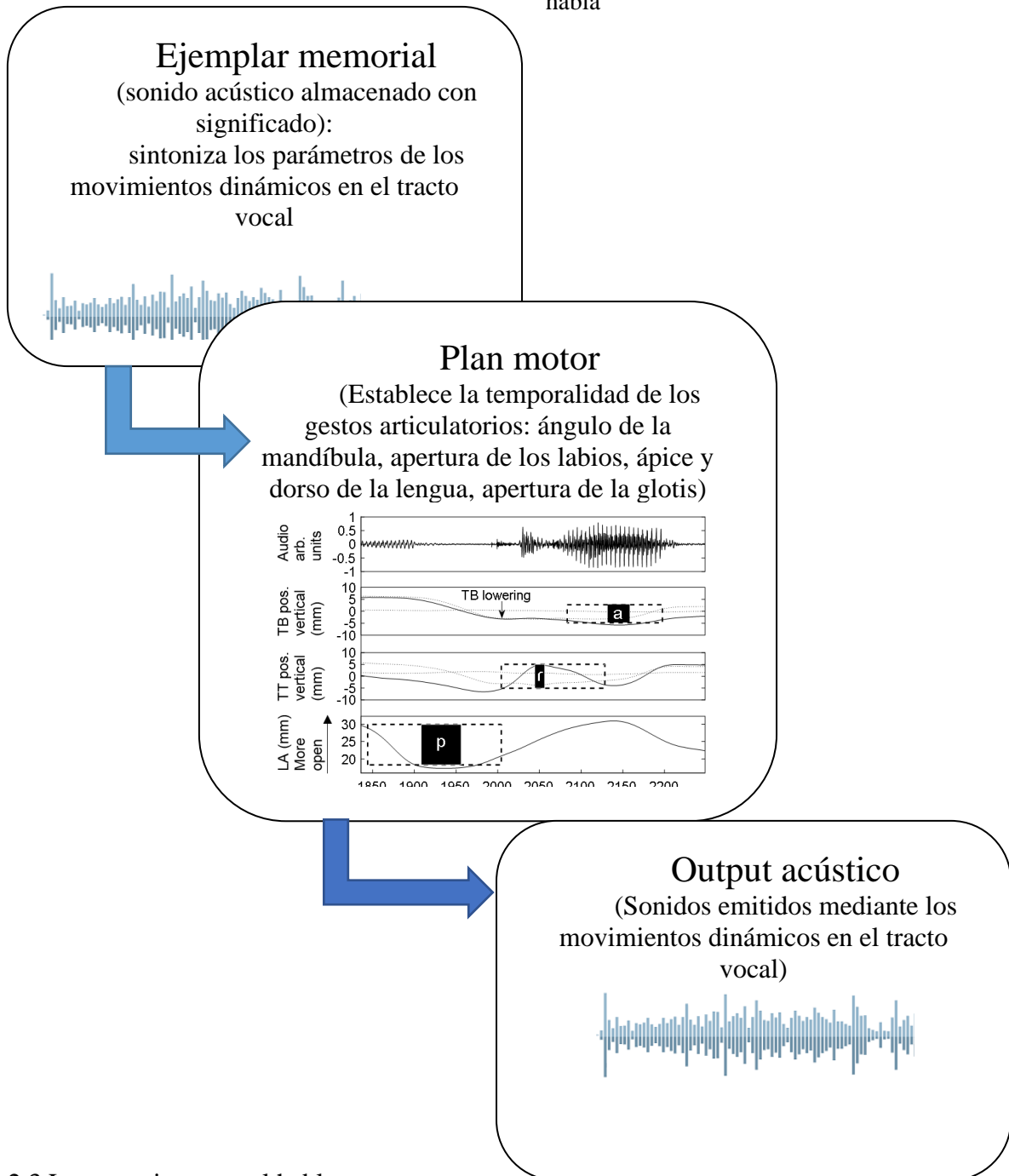
2.2 Producción del habla

El habla se produce cuando el plan motor del habla se acopla a un ejemplar almacenado en la memoria (René-Joseph Lavie). De forma que, para que una palabra se produzca, se tiene que conjugar toda la información fonológica, léxica, morfosintáctica y pragmática de la palabra previamente almacenada en la memoria con el plan motor habla, el cual actúa en diferentes partes del tracto vocal.

El plan motor del habla especifica los diferentes movimientos y la temporalidad que se debe llevar a cabo para la articulación de los sonidos. Este plan se encarga de la coordinación de los diferentes órganos del tracto vocal, desde la glotis hasta las cavidades orales y nasales. Este proceso comienza en el sistema respiratorio donde se equilibra el flujo del aire.

El aire llega al tracto vocal compuesto principalmente por la laringe, la cual contiene las cuerdas vocales y la glotis. Si la glotis está cerrada, el aire pasa a través produciendo vibración de las cuerdas vocales pero, si la glotis está abierta, no hay vibración. Esto ocasiona una fluctuación de la presión del aire que corresponde a la frecuencia más baja de una onda periódica conocida como frecuencia fundamental, F0 (Johnstone 10).

Figura 1. Elaboración propia. Ejemplificación de los pasos en la producción del habla



2.3 Las emociones en el habla

El modelo fuente-filtro de la producción del habla ayuda a entender cómo la voz facilita información sobre el estado emocional de una persona (Bachorowski 54). Este modelo establece que la producción del habla sucede a través de dos fases: una fuente y un filtro. La fuente del sonido se origina en la glotis y las cavidades subglóticas, después pasa a los filtros que son las cavidades bucales y nasales.

Relacionado con la fuente del sonido, el parámetro acústico que más se ha analizado en estudios sobre las emociones en la voz (Scherer; Abelin/Allwood; Johnstone;

Bachorowski/Owren; Yildirim/Murtaza/Lee) es la frecuencia fundamental — F0 — la cual “corresponde a la frecuencia en la abertura y cierre de los pliegues vocales. A lo largo del tiempo, este parámetro se puede entender como una curva melódica” (Llisterri). En un estudio (Scherer/Banse/Wallbot/Goldbeck), por ejemplo, se examinaron las características acústicas de estímulos neutros y afectivos realizados por cuatro actores. Los resultados mostraron que miedo y alegría están asociados con una alta frecuencia fundamental al ser comparados con estímulos neutros. En otro estudio (Bachorowski), gracias a un análisis de la frecuencia fundamental, concluyó que la tristeza suele presentar una baja frecuencia fundamental.

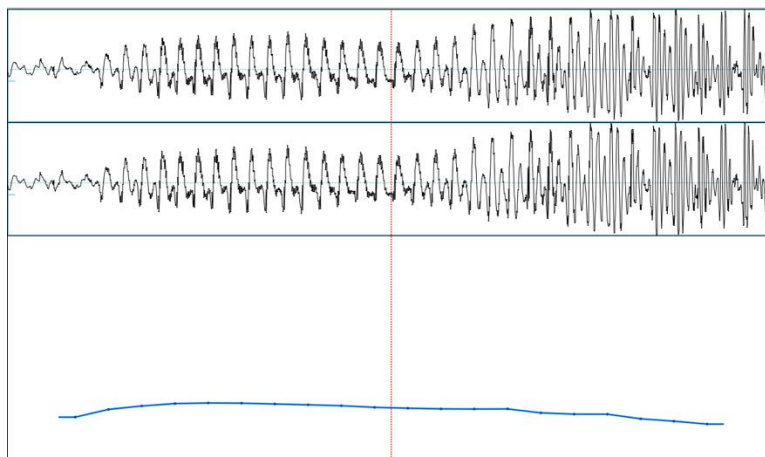
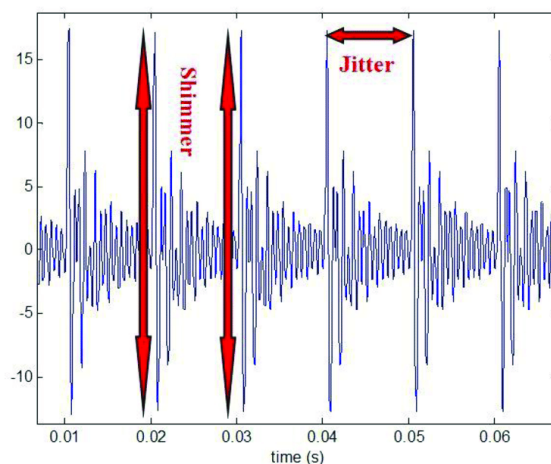


Figura 2. Elaboración propia. Muestra las variaciones de la frecuencia fundamental a través de la curva melódica. La línea azul corresponde al timbre de la voz.

Respecto al timbre de la voz, diversos estudios han concluido que el timbre presenta amplias fluctuaciones relacionadas con las emociones (Tartter 23; Laukka). El timbre aumenta con el enojo, miedo y felicidad; y decrece en emociones menos activas como la pena y la indiferencia (Tartter 24).

Además, se ha demostrado que la velocidad del habla —*speech rate*— también es un parámetro significativo y se comporta de manera similar al timbre de la voz (Abelin/Allwood). Este parámetro se calcula mediante la división entre el tiempo total del enunciado y el número de sílabas de este. Este parámetro está relacionado con la intensidad de los sonidos, es decir, la potencia con la que se emite un sonido. La velocidad del habla y la intensidad pueden aumentar en emociones activas como la felicidad y disminuir en emociones pasivas como la tristeza. Estos parámetros también parecen ser relevantes para los juicios emitidos por los oyentes (Bachorowski).

De igual forma, *jitter* y *shimmer* son parámetros relacionados con la fuente del sonido. Estos corresponden a la variabilidad y frecuencia con la que vibran las cuerdas vocales. El *jitter* es “un índice que sirve para medir la regularidad en la frecuencia donde unas ondas puede ser más anchas que otras” (Polo). El *shimmer* es una irregularidad en “la amplitud, el volumen o la intensidad, con la que se produce una onda” (Polo). Por lo tanto, en los sonidos existe cierto porcentaje de variabilidad, pero emociones como el miedo, enojo, tristeza y nerviosismo tienen repercusiones en estos parámetros.

Figura 3. Ejemplo de *shimmer* y *jitter*. Elaboración de Teixeira, J. y A. Goncalves

Los parámetros relacionados con el filtro del sonido pueden ser importantes en el estudio de las emociones en la voz, debido a que las expresiones faciales, como la posición de los labios, pueden tener efectos en el filtro del sonido. Este tipo de diferencias acústicas se reflejan en los formantes, los cuales se entienden como resonancias en el conducto vocal que corresponden a la amplitud de las frecuencias (Bachorowski; Gil 148). Es decir, los formantes son frecuencias naturales que se pueden reforzar dependiendo del tamaño y forma del tracto vocal. Por ejemplo, el formante 3 (F3) abarca los laterales del conducto vocal y, por lo tanto, F3 puede presentar mayor concentración de energía cuando una persona sonríe.

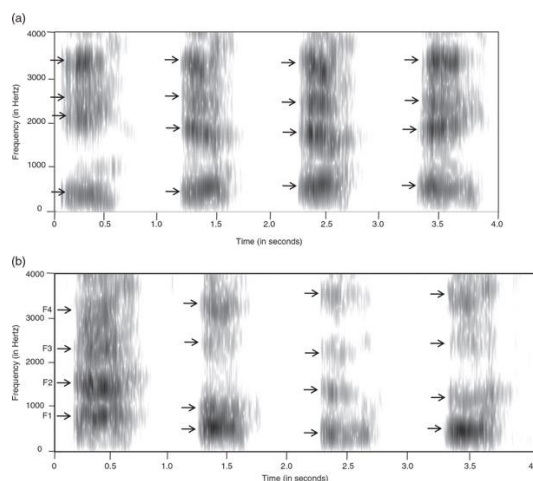


Figura 4. Los formantes en un espectrograma de Praat. Elaboración de Smith.

La mayoría de los estudios (Laukka; Johnstone; Abreu/Mathon/Mosca) identifican dos categorías de emociones a través de estos parámetros acústicos. La primera se caracteriza por una frecuencia fundamental alta, velocidad del habla alta y un timbre alto. En esta categoría se incluyen emociones relacionadas con altos niveles de excitación como el enojo y la felicidad (Abreu/Mathon/Mosca 63). La segunda categoría presenta bajos niveles de excitación y tiene como características principales una baja frecuencia fundamental y una velocidad lenta. Ejemplo de estas emociones son el aburrimiento, la indiferencia y tristeza (Abreu/Mathon/Mosca 63).

Existen tres metodologías comunes para analizar los efectos de las emociones en la voz (Abreu/Mathon/Mosca). La primera utiliza estímulos auditivos realizados por actores, lo que acarrea muchas ventajas en el análisis acústico, ya que es posible preparar estas grabaciones para que las emociones buscadas sean perceptibles en los audios. Un problema que puede surgir es la credibilidad de las grabaciones, debido a que estas están siendo manipuladas y, además, las emociones imitadas no tienen el mismo origen neuroquímico ni fisiológico que las emociones reales.

La segunda metodología emplea estímulos afectivos recolectados en el laboratorio sin el conocimiento de los participantes. En este caso se presentan una serie de estímulos con el fin de generar ciertas emociones en los sujetos. Estas pruebas están sujetas a factores externos, y pueden no generar la misma reacción en los individuos. El corpus puede ser limitado.

Por últimos, algunos estudios manejan emociones del habla espontánea donde el objetivo es analizar emociones reales. Una desventaja es que no se puede tener bajo control todas las variables del experimento y los estados emocionales son definidos en el corpus después de la recolección de las grabaciones con base en un análisis acústico. Esto conlleva un componente subjetivo, lo que significa que el investigador necesita determinar el tipo de emoción que se transmite pero, para contrarrestar la subjetividad de la selección, es necesario un análisis acústico para establecer los diferentes parámetros.

En este proyecto nos vamos a enfocar en la tercera metodología, debido a que es una de las menos estudiadas, y consideramos que las emociones espontáneas acarrearán el ímpetu psicológico que se materializa al expresar una emoción real.

2.4 Percepción de las emociones en el habla

El estudio de la percepción de las emociones en el habla ha ganado interés a la par que la investigación en la producción de las emociones en la voz, ya que la percepción de las emociones sirve de complemento a un análisis acústico. Diversos estudios (Bachorowski; Scherer; Laukka; Elfenbien/Ambady) han demostrado que los participantes son capaces de diferenciar significativamente distintas emociones, e incluso esto es posible en contextos multilingüísticos, lo que sugiere que la habilidad de detectar emociones en la voz es un componente fundamental en la comunicación humana (Bachorowski 55). Sin embargo, estudios también han indicado que la lengua materna tiene un papel importante, ya que el reconocimiento de emociones se facilita cuando se expresa en la lengua nativa (Chong/Kim/Davis).

Cuando se utilizan estímulos auditivos realizados de forma espontánea es importante notar que el porcentaje de errores es de un 60% mientras que con estímulos realizados por actores es de tan solo 40% (Scherer). Esta gran diferencia puede deberse a que en una conversación espontánea las emociones no son identificables tan fácilmente como en estímulos realizados por actores donde el foco está en transmitir las emociones de manera clara y reconocible.

3. PROYECTO 1

El propósito de este proyecto es clasificar las características acústicas de tres emociones —estado neutro, felicidad y tristeza— mediante un análisis acústico y después investigar si estos parámetros son percibidos en un test de percepción. Para esto decidimos utilizar un corpus compuesto por estímulos auditivos espontáneos con el fin de asegurar la transmisión de emociones auténticas.

Con el fin de disminuir el componente de subjetividad en la selección del corpus, decidimos trabajar conjuntamente con el Dr. Adrián Cano y la Dra. Raquel Martín del Departamento de Psiquiatría y Psicología de la Clínica de la Universidad de Navarra, donde se lleva a cabo una terapia de parejas que consiste en una entrevista de dos horas dividida en tres bloques diseñados para provocar distintas emociones. En la primera parte se tratan los aspectos cotidianos de la pareja y la emoción predominante es el estado neutro; en la segunda parte, el psiquiatra conduce la entrevista de forma que la pareja habla de los primeros meses de la relación o de los hijos con el fin de que predomine la felicidad; por último, se tratan los problemas de la relación y predomina la tristeza. Previamente la pareja ha accedido a participar y ha firmado una hoja de consentimiento aceptando que la grabación se utilice con fines científicos.

El corpus se recolectó basándose en una de estas entrevistas con el objetivo de asegurar la presencia de las emociones buscadas. Esta entrevista tenía una duración de 123 minutos y aparecían tres sujetos: el psiquiatra, quien se encargaba de realizar las preguntas y una pareja, hombre y mujer de mediana edad procedentes de Valencia.

El corpus consiste de 180 estímulos auditivos, de los 60 se enfocan en un estado emocional neutro, otros 60 se enfocan en la felicidad y los últimos, en la tristeza. En esta selección se intentó escoger palabras donde la carga semántica fuese lo más neutral posible, con el fin de que la carga emocional estuviese principalmente focalizada en el aspecto acústico. Sin embargo, las emociones son transmitidas por diversos medios, así que, al prescindir de un aspecto, la percepción de la emoción puede cambiar. Por ejemplo, es común que los hablantes usen determinados marcadores discursivos al transmitir emociones, como interjecciones, así como “¡ah!” en caso de sorpresa. Por lo tanto, nuestro corpus se compone de muchos estímulos que pueden tener una valencia semántica positiva o negativa.

3.1 Experimentos: análisis acústico y percepción acústica

En este proyecto, llevamos a cabo dos experimentos: el primero consiste en un análisis acústico de estímulos afectivos y el segundo se enfoca en la percepción acústica de estos estímulos. El análisis acústico se realizó a través del software Praat, versión 6.7.37. El corpus fue segmentado usando WebMAUS (Kisler/Reichel/Schiel 2017). La extracción de las claves acústicas se realizó de forma manual utilizando las funciones de Praat. El *jitter*, *shimmer* y la frecuencia fundamental se extrajeron con base en las vocales de los estímulos. La velocidad del habla se calculó mediante la división entre el tiempo total de cada palabra y el número de sílabas de esta.

En el experimento de percepción acústica utilizamos el software E-Prime. 180 estímulos acústicos se presentaron de manera aleatoria a 30 hispanohablantes y 10

anglohablantes. Los hablantes ingleses tenían poco o nulo conocimiento de español. Los participantes tenían la tarea de escuchar cada estímulo y señalar la emoción percibida.

3.2 Resultados: análisis acústico

El análisis se dividió con base en la emoción y el género. En la tabla 1 se puede apreciar la media de todas las variables examinadas.

Emoción	Género	Intensidad (dB)	<i>Speech rate</i>	F0 (Hz)	F3 (Hz)	<i>Jitter</i>	<i>Shimmer</i>
neutral	hombre	77,56	0,19	93, 8321	2630,79	2,60%	11,50%
neutral	mujer	75,99	0.16	170,25	2727,35	2,68%	12,33%
feliz	hombre	78, 36	0,21	100,33	2723,29	1,28%	11,52%
feliz	mujer	78,21	0.16	203,20	2811,13	1,79%	11,34%
triste	hombre	76,05	0,20	98,06	2958,03	3,51%	13,55%
triste	mujer	77,85	0,15	181,03	2747,52	3,12%	15,94%

Tabla 1. Resultados del análisis acústico

Para el análisis acústico se llevo a cabo una prueba unilateral t en R Studio. Esta prueba permite entender si las diferencias entre los distintos parámetros pueden haber ocurrido por casualidad o si existe una diferencia estadísticamente significativa.

Respecto a los estímulos del hombre, el análisis demostró que hay una diferencia significativa en la velocidad del habla al comparar estímulos neutros y estímulos felices, $t(41) = 1.49$, $p < .05$. También, se encontró diferencia significativa en el parámetro de F3 al comparar estímulos neutros y tristes, $t(49) = 3.05$, $p < .05$, y al comparar estímulos felices y tristes, $t(48) = 1.96$, $p < .05$. La intensidad del habla también demostró resultados significativos, respecto a estímulos neutros y tristes, $t(49) = 1.38$, $p < .05$, y estímulos felices y tristes, $t(48) = 1.93$, $p < .05$. El análisis mostró diferencia significativa en la velocidad del habla entre estímulos felices y neutros, $t(41) = 1.49$, $p < .05$.

Respecto a los estímulos de la mujer, el análisis mostró diferencia significativa en la frecuencia fundamental en los estímulos neutros y felices, $t(49) = 2.87$, $p < .05$, y en los estímulos felices y tristes, $t(54) = 1.37$, $p < .05$. F3 también demostró diferencia significativa en los mismos estímulos felices y neutros, $t(48) = 1.93$, $p < .05$ y en estímulos felices y tristes, $t(53) = 1.51$, $p < .05$. En la velocidad del habla también se encontró diferencia significativa en estímulos negativos y positivos, $t(54) = 2.20$, $p < .05$. Por último, hay diferencia significativa en la intensidad de estímulos neutros y positivos, $t(44) = 1.71$, $p < .05$, y en estímulos neutros y negativos, $t(48) = 1.56$, $p < .05$.

3.2 Resultados: percepción acústica

Los resultados del experimento de percepción acústica se analizaron con la prueba unilateral t y un ANOVA —*Analysis of Variance*— en R Studio. El ANOVA permite evaluar simultáneamente el efecto de la agrupación de dos variables. Los hispanoparlantes obtuvieron un 40% en exactitud mientras que los anglohablantes un 39%. La prueba t no demostró diferencia significativa entre ambos grupos lingüísticos, $t(30)=1.22, p>.05$.

Las siguientes gráficas ilustran los resultados obtenidos en las tres emociones analizadas. La prueba unilateral ANOVA no determinó diferencias significativas entre las medias de los grupos inglés y español en los estímulos neutros ($F(1,29)=0.934, p=0.1$); tampoco en los estímulos felices ($F(1,29)=0.026, p=0.1$) ni en los tristes ($F(1,29)=0.45, p=0.1$).

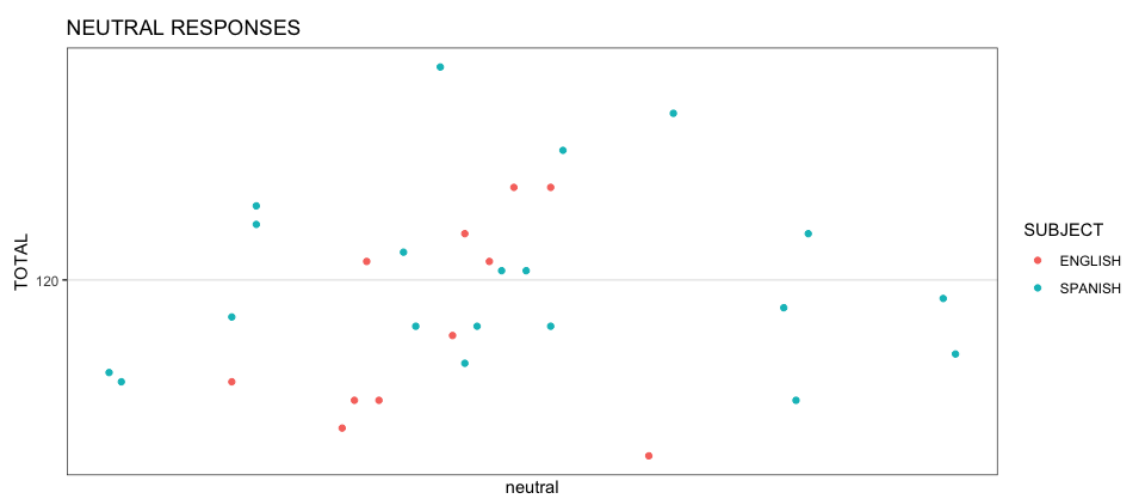


Figura 5. Gráfico de dispersión donde se muestra la relación entre las respuestas de los sujetos anglohablantes e hispanohablantes a los estímulos neutrales, “x” se refiere a la emoción neutral y “y” al número total de respuestas.

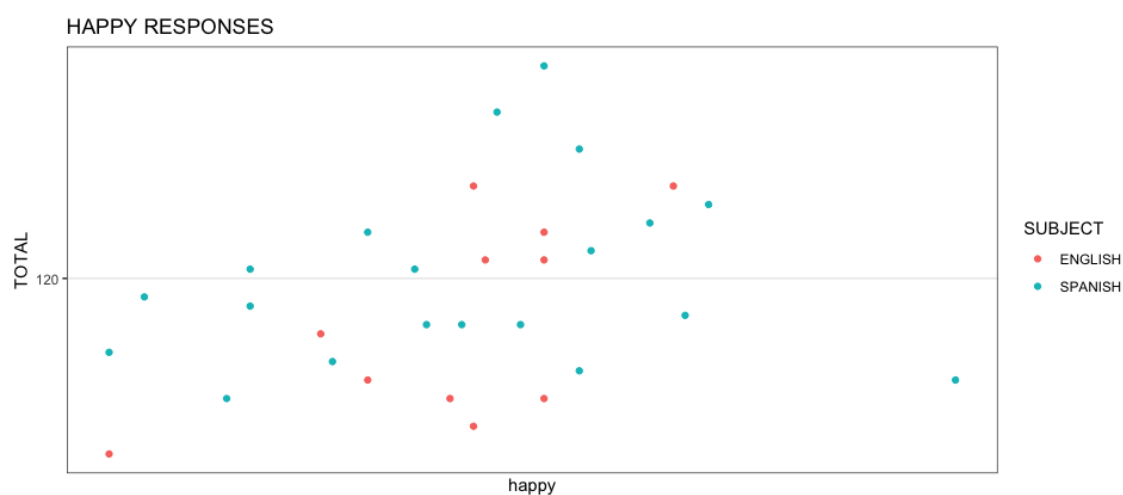


Figura 6. Gráfico de dispersión donde se muestra la relación entre las respuestas de los sujetos anglohablantes e hispanohablantes a los estímulos felices.

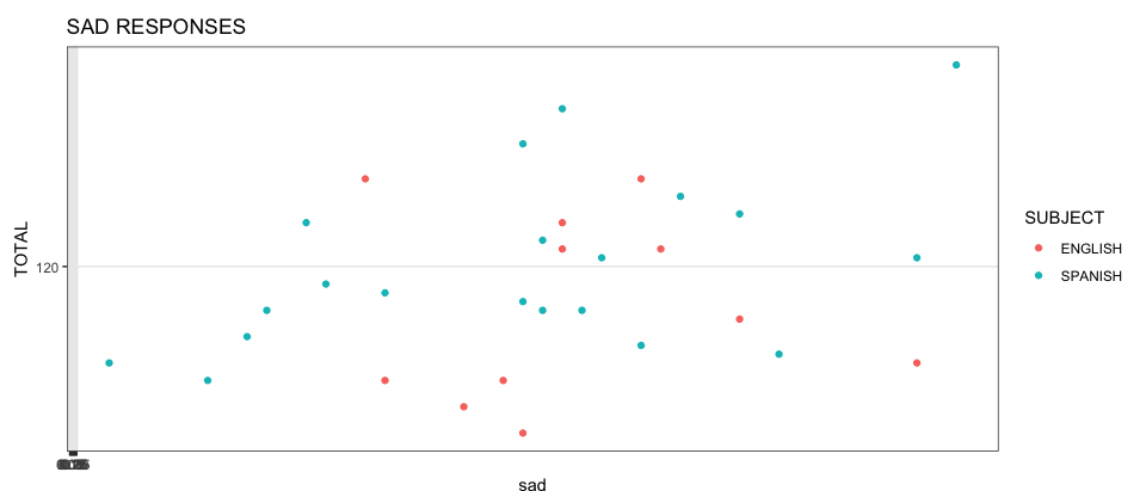


Figura 7. Gráfico de dispersión donde se muestra la relación entre las respuestas de los sujetos anglohablantes e hispanohablantes a los estímulos neutrales.

Los hispanohablantes reconocieron el 56% de los estímulos neutros, 47.62% de los felices y 18.27% de los estímulos tristes. La precisión en los anglohablantes fue de 56% en el estado neutro, 39% en felicidad y 25% en tristeza. En general, al identificar el estado neutro las mujeres de ambos grupos lingüísticos obtuvieron menos errores que los hombres, el cual se reconoció con un 60% de precisión, la emoción feliz se reconoció un 43% y la tristeza tan solo un 25%.

3.3 Conclusión y discusión

De forma que la frecuencia fundamental, F3, la velocidad del habla y la intensidad son parámetros acústicos que pueden señalar los diferentes estados emocionales y estos pueden ayudar a los oyentes a identificar emociones en la voz. El habla feliz es producida con una mayor frecuencia fundamental, intensidad y velocidad que el habla triste. El habla triste se caracteriza por tener un porcentaje alto de *jitter* y *shimmer* comparado con el habla neutral y feliz.

En el test de percepción, el estado neutro fue reconocido con mayor facilidad que la felicidad y tristeza, y no se encontró diferencia significativa entre hablantes nativos y no nativos. Esto cuestiona la idea de que los hablantes identifican con mayor precisión las emociones en la voz cuando se expresan en su lengua materna (Abelin/Allwood). La razón puede ser que los hablantes nativos no solamente tenían en cuenta la modalidad acústica, sino que estaban influenciados por el significado de las palabras, y la presencia de estímulos ambiguos puede haber generado confusión. Por ejemplo, los hispanohablantes escogieron felicidad en palabras como “playa”, la cual previamente se había clasificado con parámetros acústicos relacionados con la tristeza. En el caso de los anglohablantes esta confusión no se encontró, lo que sugiere que, al no entender el significado de las palabras, solo podían enfocarse en la modalidad acústica.

Además, a la hora de percibir estímulos acústicos es importante tener en cuenta la personalidad del sujeto que produce el estímulo, ya que muchas personas son más

expresivas que otras, lo cual se puede reflejar en la detección de las emociones (Laukka). Esto podría ser la razón por la que los oyentes identificaron mejor los estímulos de la mujer que los del hombre. Futuros estudios deberían intentar analizar este tema más en detalle y sistemáticamente.

4. PROYECTO II

De acuerdo con los resultados obtenidos en el primer proyecto, decidimos estudiar la interacción entre las modalidades auditiva y semántica al percibir estímulos afectivos, utilizando el corpus del primer proyecto. Para esto empleamos una técnica psicolingüística conocida como *priming*. En específico, *masked priming*.

Este trabajo analiza la integración multisensorial de información auditiva y visual en hablantes hispanoparlantes y anglohablantes. Este estudio aborda cómo se integran las modalidades auditiva y semántica cuando se percibe el habla emocional (González/Gibson). El objetivo es entender hasta qué punto la respuesta consciente a un estímulo auditivo afectivo puede estar influenciada por una palabra afectiva anterior presentada de manera visual.

4.1 *Masked priming* en el lenguaje

Priming es una técnica de percepción visual donde dos estímulos son presentados sucesivamente. El primer estímulo se conoce como *prime* y el segundo como *target*. *Priming* ocurre cuando el *prime* facilita la respuesta o identificación del objetivo, en relación con un estado neutro (Forster/Davis). En *masked priming* la diferencia es que el *prime* aparece por un tiempo tan corto que los sujetos son en gran parte inconscientes de la presencia de este estímulo (Forster/Mahon/Hector).

La mayoría de los estudios siguen el siguiente método: un *prime* se presenta entre 35 y 50 milisegundos y este es subsecuentemente remplazado por un *target*, que es perfectamente perceptible (Davis/Kim 2015). Para lograr esto es importante la presencia de una máscara (*mask*) introducida inmediatamente antes que el *prime* y otra después de este (*backward mask*), así, el *target* es perfectamente visible. Por lo tanto, la secuencia es: *mask-prime-backward mask-target* (Forster/Davis).

Generalmente, esta técnica es utilizada para examinar los efectos automáticos del reconocimiento visual de una palabra, es decir, la asociación semántica de palabras. Un estudio, por ejemplo, demostró que la palabra “NURSE”—*target*— se identificó más rápidamente cuando precedía a un *prime* semánticamente relacionado como “doctor”, que cuando precedía a otro *prime* con el que no guardaba ninguna relación, tal como “butter”(Bodner/Masson).

De forma que, en esta técnica un estímulo visible se procesa con mayor eficacia —más rápido y con menos errores— cuando precede a un *prime* con el que guarda alguna relación que cuando precede a un *prime* con el que no tiene relación (Davis/Kim 2015). Esta facilitación, conocida como *priming* positivo, puede actuar como un sondeo de la estructura léxica, debido a que se revela la presencia de vínculos entre el *prime* y el *target* que se desencadenan sin la necesidad de una deliberación consciente por parte del sujeto.

4.2 Experimento

Se examinaron las respuestas de 12 hispanohablantes y 12 anglohablantes con poco o nulo conocimiento de español. Durante el experimento principal a los participantes se les pidió enfocarse en el estímulo acústico y señalar si percibían tristeza o felicidad. Después del experimento principal, los sujetos llevaron a cabo una prueba de visibilidad del *prime* para comprobar qué tan visible es el *prime*. En esta prueba se le explicó a los participantes la presencia del *prime*.

Es importante tener en cuenta el cambio de instrucciones en el experimento principal y la prueba de visibilidad, debido a que, aunque es necesario que los participantes no puedan identificar el *prime* mejor que la casualidad, al mostrarles la presencia del *prime*, no se puede evaluar objetivamente si ellos fueron conscientes del *prime* en el experimento principal. Si los participantes llegaran a tener más del 50% de respuestas correctas en esta prueba, esto significaría que la técnica del *masked priming* no está cumpliendo su propósito: analizar la influencia y relación entre un estímulo procesado de manera inconsciente y un segundo estímulo procesado conscientemente.

4.3 Metodología

En este proyecto el *target* corresponde a los estímulos afectivos utilizados en el primer proyecto, con la diferencia de que en este experimento prescindimos de los estímulos neutros y solo presentamos los de felicidad y tristeza. El *prime* corresponde a una palabra semántica —positiva o negativa— y aparece por tan solo 45 milisegundos. En el grupo de hispanohablantes el *prime* es “feliz” o “triste”, y “smile” o “sorrow” para los anglohablantes. Para enmascarar el *prime* utilizamos una serie de símbolos, que han sido utilizados en previos experimentos: “#&#&#&#&#&#” (Davis/Kim 2011). El diseño del experimento se realizó a través del software DMDX (Forster/Forster).

El experimento, por lo tanto, consta de dos estímulos, *prime* y *target*. El *prime* es una palabra semántica que puede ser positiva o negativa —feliz o triste—. El *target* es un estímulo acústico que puede ser positivo o negativo. De manera aleatoria los estímulos van a ser congruentes o incongruentes, es decir, cuando sean congruentes el *prime* tendrá la misma carga afectiva que el *target*. Por lo tanto, las condiciones congruentes son: “feliz” + estímulo auditivo feliz y “triste” + estímulo auditivo triste. Las condiciones incongruentes son: “feliz” + estímulo auditivo triste y “triste” + estímulo auditivo feliz.

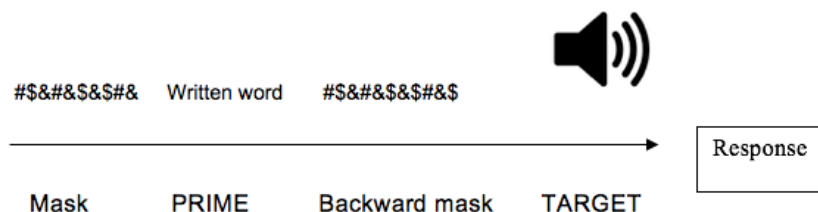


Figura 8. Diseño del experimento.

4.4 Análisis estadístico

Para el análisis utilizamos una serie de modelos lineales de efectos mixtos mediante el uso del paquete “lme4” en R Studio. El modelo de efectos mixtos se prefirió en lugar de un ANOVA para evitar la inflación de la proporción de F (Gibson/Sotiropoulou/Tobin/Gafos 9). La tabla 2 enlista las variables dependientes, el predictor, y los efectos aleatorios estudiados en el análisis. En todos los casos, se analizaron los efectos principales y sus interacciones. Los participantes fueron considerados como un efecto aleatorio, debido a que esperamos diferencias entre sujetos que no se relacionan con la influencia de los predictores.

VARIABLES DEPENDIENTES	Descripción
Idioma	Inglés/español
Tipo de <i>prime</i>	Positivo/Negativo
Tiempo de reacción (RT)	ms
Porcentaje de error	%
PREDICTOR	Descripción
Condición	COND_1: incongruente(prime positivo y audio negativo) COND_2: incongruente (prime negativo y audio positivo) COND_3: congruente (prime positivo y audio positivo) COND_4: congruente (prime negativo y audio negativo)
EFFECTOS ALEATORIOS	Descripción
Participantes	12 anglohablantes 12 hispanohablaantes

Tabla 2. Esta tabla muestra las variables analizadas.

4.5 Resultados

Los resultados de la prueba de visibilidad del *prime* mostraron que el rendimiento no fue superior al azar. Por lo tanto, suponemos que el *prime* fue procesado inconscientemente por los participantes durante el experimento principal (González/Gibson). Los resultados generales se presentan en la tabla 3:

	Tipo de <i>prime</i>	Congruente		Incongruente	
		RT (ms)	%ERROR	RT(ms)	%ERROR
ESPAÑOL	<i>Prime</i> positivo(feliz)	1089	61%	1118	72%
	<i>Prime</i> negativo(triste)	1091	76%	1117	78%
INGLÉS	<i>Prime</i> positivo(feliz)	964	39%	1002	51%
	<i>Prime</i> negativo(triste)	870	50%	966	52%

Tabla 3. Tabla que expone los resultados del tiempo de reacción (ms) y el porcentaje de error obtenido en cada condición de los hispanoparlantes y angloparlantes.

4.5.1 Resultados hispanohablantes

Los resultados demostraron que el porcentaje de error es menor en los *primes* congruentes positivos comparado con los *primes* incongruentes positivos, pero no se da una gran diferencia en el caso de los *primes* congruentes negativos. Se encontraron efectos significativos en el porcentaje de error en las diferentes condiciones ($\chi^2[1, n=7] = 6.77, p < 0.01$). Pero no se encontraron efectos robustos en el tiempo de reacción ($\chi^2[1, n=7] = 0, p < 1$). Las figuras 9 y 10 ilustran estas diferencias.

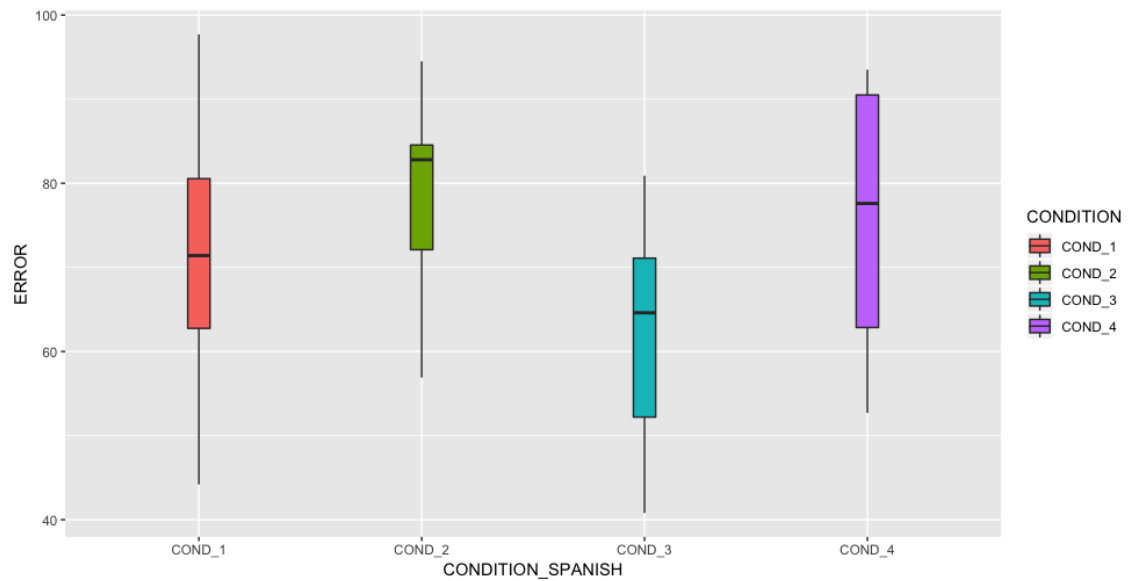


Figura 9. Este diagrama de caja muestra el porcentaje de error de las condiciones incongruentes con *prime* positivo (COND_1) y con *prime* negativo (COND_2); y congruentes con *prime* positivo (COND_3) y con *prime* negativo (COND_4)

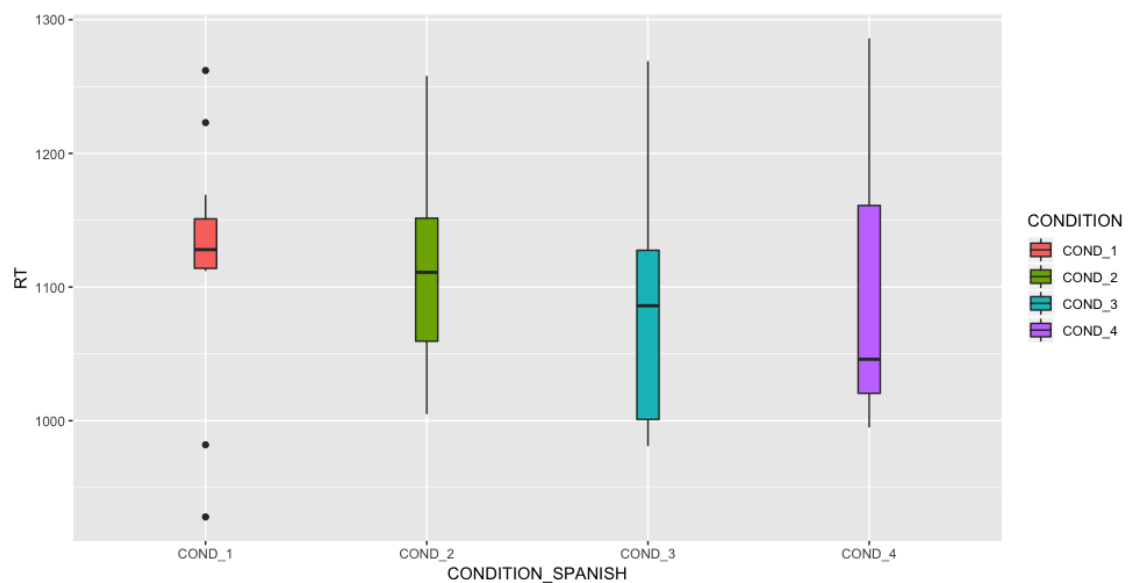


Figura 10. Este diagrama de caja muestra el tiempo de reacción (ms) de las condiciones incongruentes con *prime* positivo (COND_1) y con *prime*

negativo(COND_2); y congruentes con *prime* positivo(COND_3) y con *prime* negativo(COND_4)

4.5.2 Resultados anglohablantes

Los resultados demostraron, tal como se esperaba, que el porcentaje de error es menor en los *primes* congruentes que en los incongruentes, sobre todo aquellos que son positivos. La figura 11 ilustra las diferencias en el porcentaje de error de las diferentes condiciones. Efectos robustos se encontraron en el porcentaje de error respecto a las cuatro condiciones ($\chi^2[1, n = 11] = 35.53, p < 0.001$). Pero no se encontró en el tiempo de reacción, ($\chi^2[1, n = 11] = 3.75, p < 0.1$). La figura 12 muestra las diferencias en los tiempo de reacción de cada condición.

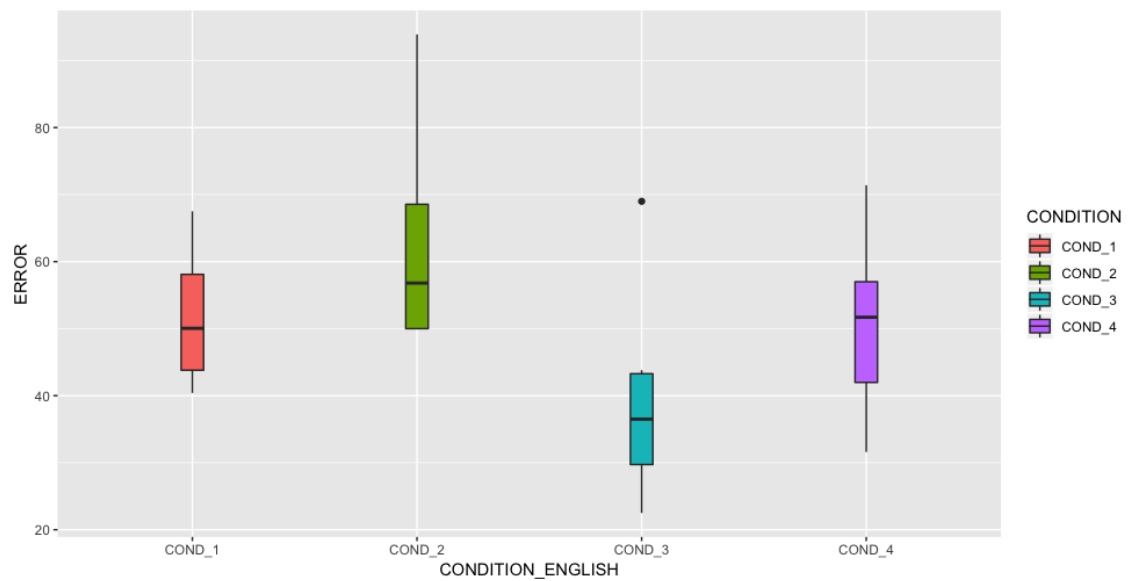


Figura 11. Este diagrama de caja muestra el porcentaje de error de las condiciones incongruentes con *prime* positivo (COND_1) y con *prime* negativo(COND_2); y congruentes con *prime* positivo(COND_3) y con *prime* negativo(COND_4).

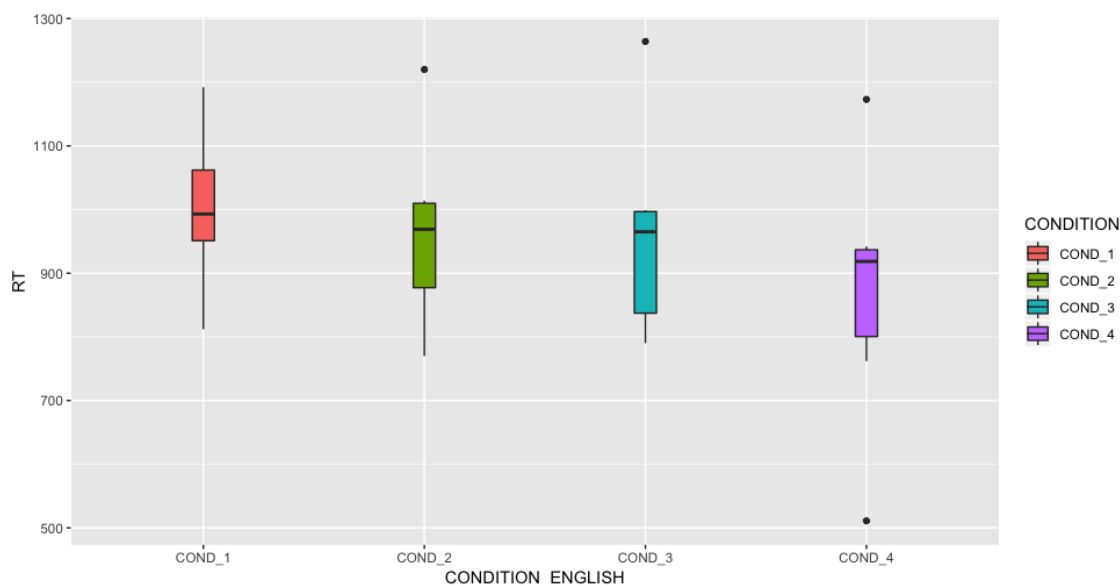


Figura 12. Este diagrama de caja muestra el tiempo de reacción (ms) de las condiciones incongruentes con *prime* positivo (COND_1) y con *prime* negativo (COND_2); y congruentes con *prime* positivo (COND_3) y con *prime* negativo (COND_4).

Los resultados demostraron una relación entre el tipo de condición y el porcentaje de error, siendo este inferior cuando se presentaba una condición congruente, sobre todo cuando era positiva. Además, aunque no se encontraron efectos robustos en el tiempo de reacción en las distintas condiciones, los resultados muestran una tendencia a responder con mayor rapidez cuando se presenta una condición congruente.

Estos resultados también se compararon con los resultados obtenidos en el primer proyecto. En el caso de los participantes ingleses, el primer proyecto reveló que el porcentaje de error en estímulos felices fue de 61% mientras que en este proyecto el porcentaje de error fue de tan solo 39%. Respecto a los estímulos tristes, en el primer proyecto el porcentaje de error fue de 75% y en este proyecto, fue de 50%.

4.6 Conclusión

Los resultados apuntan a que, en el caso de los anglohablantes, la técnica *masked priming* mejora el rendimiento en la percepción de estímulos afectivos. Podemos concluir que las representaciones entre el *prime* —palabra semántica— y el *target* —estímulo auditivo— se desencadenan de forma más automática y eficiente. Estos resultados están en línea con lo concluido en Navarra/Alsus/Velasco/Soto-Faraco (88): “la percepción combinada del lenguaje visual y auditivo acelera las respuestas”.

Asimismo, al comparar estos resultados con aquellos obtenidos en el primer proyecto, en el caso de los anglohablantes, se puede concluir que la mejoría en el desempeño está relacionada con la familiaridad del lenguaje procesado. Además, los resultados sugieren que los hablantes dan más valor a las claves visuales cuando escuchan una lengua no conocida que cuando escuchan su lengua materna. Estos resultados son

similares a los encontrados por Hazan: los oyentes no nativos le dan más valor a las claves visuales que a las acústicas.

En el caso de los hispanohablantes no se encontró este vínculo, sino que más bien los resultados apuntan a que no existe una tendencia clara. Esto podría relacionarse con lo concluido por Navarra/Alsius/Velasco/Soto-Faraco (87), “en el caso de un lenguaje conocido, para que se de un efecto significativo la información visual debe aparecer antes de lo que se presenta un habla desconocida”. Por lo tanto, sería interesante investigar más en detalle la técnica *masked priming* presentando el *prime* a los hablantes nativos durante más de 45 milisegundos.

Otro factor a tener en cuenta es la adaptación de los oyentes durante el experimento, debido a que, a pesar de que el *prime* está enmascarado, conforme realizan el experimento lo perciben con mayor facilidad y, por lo tanto, el efecto puede disminuir.

Además, los resultados en el tiempo de reacción puede depender de diferentes factores, por ejemplo, en este experimento para responder se utilizaron las teclas, “Alt+teclas Mayus derecha” para felicidad y “Alt + teclas Mayus izquierda” para tristeza, de forma que al pedir a los participantes que usen las dos manos, es probable que la mano dominante responda más rápido (Mangen/Anda/ Oxborough). Para eliminar este factor, muchos estudios en lugar del teclado utilizan controles de videojuegos donde se tiene que presionar un gatillo. De modo que se tiene que considerar que, además de la percepción, el tiempo de reacción implica también medir el control motor del sujeto.

Asimismo, respecto a los dos proyectos consideramos importante que futuros estudios controlen mejor el estado emocional de los participantes, qué tanto durmieron la noche anterior a la prueba y el periodo de atención de los participantes, ya que estos factores también pueden influir en cómo se perciben ciertos estímulos (Fernández-Caballero/Martínez-Rodrigo/Pastor). Sería interesante realizar el mismo proyecto comparando estas variables.

5. RELEVANCIA Y FUTUROS ESTUDIOS

En los últimos años, gracias al auge de la inteligencia artificial, el desarrollo de asistentes automáticos del habla ha aumentado considerablemente. Así, la detección automática de emociones se ha convertido en un campo de gran valor en el aprendizaje automatizado —*machine learning*—, ya que permite a un ordenador detectar de manera automática las emociones en el habla natural. Para que esto se logre, es importante proporcionar al programa patrones relevantes. Por lo tanto, para el correcto desarrollo de estos sistemas de inteligencia artificial, se necesitan datos fundamentados en estudios experimentales.

Este trabajo de investigación contribuye en este campo, debido a que aborda dos ideas importantes. La primera idea es un enfoque en la lengua española. A pesar de que hoy en día ya existen varios programas especializados en la detección automática de las emociones, entre ellos OpenSmile o Emovoice, la gran mayoría de estos estudios están enfocados en la habla inglesa, alemana y china. Consideramos de gran importancia tratar este campo con datos de hispanohablantes para poder aportar información relevante al desarrollo de sistemas de inteligencia artificial especializados en esta lengua. Este campo aporta muchas ventajas de progreso en diversas áreas, por ejemplo, en el de la salud, ya

que estos programas pueden ayudar a la detección temprana de problemas mentales como la depresión (Kiss/ Sztahó/ Vicsi).

La segunda idea es que este trabajo aporta una visión multidisciplinaria en la producción y percepción de las emociones en el habla, debido a que se enfoca en la integración de señales visuales y auditivas. Además, entender cómo se producen y perciben las emociones involucra a muchas ciencias, como la neurología, psicología, lingüística, estadística e informática. De forma que es primordial crear puentes entre estas disciplinas para poder abarcar este estudio de la manera más precisa.

El objetivo es aportar información relevante teniendo en cuenta los procesos involucrados en la percepción y producción de emociones para diseñar programas cada vez más sofisticados. En especial, los resultados sugieren que hablantes nativos y no nativos integran de diferente manera las señales auditivas y visuales. Futuros análisis deben buscar complementar este estudio utilizando técnicas neuropsicológicas como un EEG —electroencefalografía—, ya que la integración de claves visuales y auditivas se origina a través de diferentes interacciones neuronales en distintas escalas (Wassenhove/ Grant/ Poeppel). Otra implementación interesante sería modificar la palabra semántica en el experimento de *masked priming* por expresiones faciales para entender mejor la clave sensorial que prevalece cuando se percibe un estado emocional.

6. BIBLIOGRAFÍA

- Abakumova, Olga y Hermenegildo López González. “Descripción de la percepción del habla emocional infantil: teoría y práctica”, *E.H Filología*, 30 (2008): 351-60
- Abelin, Asa y Jens Allwood. “Cross Linguistic Interpretation of Emotional Prosody”, *Proceedings of the ISCA Workshop on Speech and Emotion*, 1 (2000): 110-13.
- Abreu, Sophie, Catherine Mathon y Alessandra Mosca. “Prosodic Cues of Anger in French: A Comparative Study with Two Other Romance Languages, Italian and Portuguese”. *The Role of Prosody in Affective Speech*. Ed. Sylvie Hancil. Bern: Peter Lang AG, 2009: 61-93.
- Bachorowski, Jo-Anne. “Vocal expression and perception of emotion”, *Psychological science* 8 (1999): 53-57
- Bachorowski, Jo-Anne y Michael Owren. “Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context”. *Psychological Science* 6 (1995): 219-24
- Bodner, Glen y Michael Masson. “Beyond spreading activation: an influence of relatedness proportion on masked semantic priming”. *Psychonomic Bulletin & Review* 10.3 (2003): 645-52
- Chóliz, Mariano. *Psicología de la emoción: el proceso emocional*. 2005
- Chong, Che Seng, Jeeseun Kim y Chris Davis. “Development of an Audivisual Cantonese Emotional Speech Database”. *The 17th conference of the Oriental chapter of the International Coordinating Committee on Speech Databases and Speech I/O Systems and Assessment* (2014)
- Darwin, Charles. *La expresión de las emociones en los animales y el en el hombre*. Madrid: Alianza, 1984.

- Davis, Chris y Jeesun Kim. "Masked cross-modal priming turns on a glimpse of the prime". *Consciousness and Cognition*, 33 (2015): 457-71.
- Davis, Chris y Jeesun Kim. "What is a mask? Information forward and backward masks". *Quarterly Journal of Experimental Psychology* (2011): 1990-2002
- Davis, Chris y Jeesun Kim. "Masked speech priming: neighborhood size matters". *Acoustic Society of America* 127 (2010): 2110-13
- Eyben, Florian, Felix Weninger, Florian Gross y Björn Schuller. "Recent developments in openSMILE, the Munich Open-Source Multimedia Feature Extractor". Barcelona: *Proc. ACM Multimedia (MM)*, 2013.
- Fernández-Caballero, Antonio, Arturo Martínez-Rodrigo y José Manuel Pastor. "Smart environment architecture for emotion detection and regulation", *Journal of Biomedical Informatics* (2016): 55-73
- Forster, Kenneth y Chris Davis. "Repetition priming and frequency attenuation in lexical Access". *JEP: Learning, Memory, and Cognition*, 10.4 (1984): 680-98.
- Forster, Kenneth y Jonathan Forster. "DMDX: A Windows display program with millisecond accuracy", *Behavior Research Methods, Instruments, & Computers*, 35.1 (2003): 116-25.
- Forster, Kenneth, Kathleen Mohan y Jo Hector. "The Mechanics of Masked Priming". *Masked Priming: The State of Art*. Eds. Sachiko Kinoshita y Stephen Lupker. Nueva York: EBSCO, 2003, 2-52.
- Gibson, Mark, Stavroula Sotiropoulou, Stephen Tobin y Adamantios Gafos. "Temporal Aspects of Word Initial Single Consonants and Consonants in Clusters in Spanish" *Phonetica* (2019).
- Gil Fernández, Juana. *Los sonidos del lenguaje*, Madrid: Síntesis, 1988.
- González Machorro, Mónica y Mark Gibson. "Detecting Emotions in Speech: A Cross-Linguistic Study Using Masked Affective Priming". *Presentation at the 4th Young Female Researchers in Speech Science & Technology Workshop*. (2019).
- Hazan, Valerie. "The impact of visual training on the perception and production of a non-native phonetic contrast" *ISCA ARCHIVE* (2007).
- Hornykiewicz, Oleh. "Chemical neuroanatomy of the basal ganglia — normal and in Parkinson's disease". *Journal of chemical neuroanatomy*, 22. 3(2001): 3-12
- Izard, Carroll. *The Psychology of Emotions*. Plenum Press. 1991.
- Johnstone, Tom. *The effect of emotion on voice production and speech acoustics*. University of Western Australia, 2001.
- Kim, Jeesun y Chris Davis. "Perceiving emotion from a talker: how face and voice work together". *Visual Cognition*, 20. 8 (2012).
- Kiss, Gábor, David Sztahó y Klára Vicsi. "Depression State Assessment: Application for detection of depression by speech". *INTERSPEECH 2019: show & tell contribution* (2019): 966-67.
- Kouider, Sid y Emmanuel Dupoux. "A functional disconnection between spoken and visual word recognition: Evidence from unconscious priming". *Cognition*, 82 (2001): 35-49.
- Kouider, Sid y Emmanuel Dupoux. "Subliminal speech priming". *Psychological science*, 16 (2005): 617-25.

- Laukka, Petri. *Vocal expression of emotion. Discrete emotions and dimensional actions*, Comprehensive Dissertations from the Faculty of Social Sciences. UPPSALA, 2004.
- Llisterri, Joaquim. "Frecuencia fundamental". *Las características acústicas de los sonidos del habla*. 20 de febrero 2020: <[http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html#Frecuencia_fundamental_\(f0\)](http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/fon_acust.html#Frecuencia_fundamental_(f0))>
- Mangen, Anne, Liss Anda y Gun Oxborough. "Handwriting versus Keyboard Writing: Effect on Word Recall", *Journal of Writing Search* (2015): 299-319.
- Menezes, Caroline, Donna Erickson y Jongyhe Han. "Cross-Linguistic Cross-modality Perception of English Sad and Happy Speech". 649-52.
- Navarra, Jordi, Agn s Alsius, Ignacio Velasco y Salvador Soto-Faraco. "Perception of audiovisual speech synchrony for native and non-native language". *Brain Research*, 1323 (2010): 84-93.
- Polo, Nuria "El discurso de Navidad del rey o c mo cambia la voz". *Sottovoce* (2016): 2 de marzo 2020 <<https://sottovoce.hypotheses.org/tag/shimmer>>
- Ren -Joseph, Lavie. "Exemplar theory in linguistics: a perspective for the cognitive subject". *Hal*. 2007.
- Russell, James, Jo-Anne Bachorowski y Jos -Miguel Fern ndez-Dols. "Facial and Vocal Expressions of Emotion". *Annual Review of Psychology* 54.1 (2003): 329-49.
- Scherer, Klaus. "Expression of emotion in voice and music". *Journal of Voice*, 9.3 (1995): 235-48.
- Scherer, Klaus, Rainer Banse, Harald Wallbott y Thomas Goldbeck. "Vocal cues in emotion and decoding". *Motivation and Emotion*. 15 (1991): 123-48
- Simonyan, Kristina, Barry Horwitz y Erich D. Jarvis. "Dopamine Regulation of Human Speech and Bird Song: A Critical Review" in *Brain Lang*, 122. 3 (2012): 142-50.
- Smith, Bruce. "Sound Spectography". *SAGE Encyclopedia of Human Communication Sciences and Disorders*
- Tartter, Vivien. "Happy Talk: Perceptual and Acoustic Effects of Smiling on Speech", *Perception & Psychophysics*, 27 (1980): 24-27.
- Teixeira, Joao Paulo y Andr  Goncalves (2014): "Accuracy of Jitter and Shimmer Measurements", *Procedia Technoogy*, 16, Elsevier, pp. 1190-99.
- Thomas, Kisler, Uwe D. Reichel y Florian Schiel. "Multilingual processing of speech via web services". *Computer Speech & Language*, 45, (2017): 326-47.
- Yildirim, Serdar, Bulut Murtaza y Chul Min Lee "An acoustic study of emotions expressed in speech". *ISCA ARCHIVE* (2004).
- Wagner, Johannes y Elisabeth Andr . "The Social Signal Interpretation Framework for Real Time Signal Processing and Recognitions". *Proceedings of INTERSPEECH 2011*. (2011).
- Wassenhove, Virginie, Ken W. Grant y David Poeppel. "Electrophysiology of Auditory-Speech Visual Integration" (2003).